Understanding data

The task of building the PPDB, BPDB and VSDB began more than 25 years ago. At that time data requirements for regulatory purposes were much simpler than today. Data for many species now included in regulatory dossiers were not required and, where data was required, often just one parameter was needed whereas today many different parameters are necessary. Consequently, our datasets are often not 'pure' - meaning that data for every active substance is not necessarily the same in terms of the specific parameter, endpoints and measurement units. This is particularly the case for our ecotoxicity data.

To harmonise the datasets as best we can we start with a 'preferred' parameter which defines the species, endpoint and metric (e.g. for temperate fish species our preferred data is the Acute 96hr LC_{50} mg l^{-1} for either *Oncorhynchus mykiss*, *Pimephales promela* or *Lepomis macrochirus*). Where we have a choice, this is our preferred data for entry into the PPDB. However, sometimes this exact option may not be available and so the 'next best' available data is used as a surrogate. This might be, for example, a different species or test length.

Where the data is not a good match for the preferred parameter information on the actual data description is given either in the data descriptor column on the left-hand side of the online version on in a description field if using the MS Access or Excel versions. In addition, the Quality Barometer data may be reduced to indicate uncertainty.

When using PPDB data, it is very important that the data description information is used in conjunction with the actual quantitative data. For older data there may be a paucity of support information.